Asymptotic normality and efficiency of the maximum likelihood estimator for the parameter of a ballistic random walk in a random environment

Mikael FALCONNET* Dasha LOUKIANOVA† Catherine MATIAS*

February 5, 2013

Abstract

We consider a one dimensional ballistic random walk evolving in a parametric independent and identically distributed random environment. We study the asymptotic properties of the maximum likelihood estimator of the parameter based on a single observation of the path till the time it reaches a distant site. We prove an asymptotic normality result for this consistent estimator as the distant site tends to infinity and establish that it achieves the Cramér-Rao bound. We also explore in a simulation setting the numerical behaviour of asymptotic confidence regions for the parameter value.

Key words : Asymptotic normality, Ballistic random walk, Confidence regions, Cramér-Rao efficiency, Maximum likelihood estimation, Random walk in random environment. *MSC 2000* : Primary 62M05, 62F12; secondary 60J25.

1 Introduction

Random walks in random environments (RWRE) are stochastic models that allow two kinds of uncertainty in physical systems: the first one is due to the heterogeneity of the environment, and the second one to the evolution of a particle in a given environment. The first studies of one-dimensional RWRE were done by Chernov (1967) with a model of DNA replication, and by Temkin (1972) in

*Laboratoire Statistique et Génome, Université d'Évry Val d'Essonne, UMR CNRS 8071, USC INRA, E-mail: {mikael.falconnet, catherine.matias}@genopole.cnrs.fr; † Laboratoire Analyse et Probabilités, Université d'Évry Val d'Essonne, E-mail: dasha.loukianova@univ-evry.fr

the field of metallurgy. From the latter work, the random media literature inherited some famous terminology such as *annealed* or *quenched* law. The limiting behaviour of the particle in Temkin's model was successively investigated by Kozlov (1973); Solomon (1975) and Kesten et al. (1975). Since these pioneer works on one-dimensional RWRE, the related literature in physics and probability theory has become richer and source of fine probabilistic results that the reader may find in recent surveys including Hughes (1996) and Zeitouni (2004).

The present work deals with the one-dimensional RWRE where we investigate a different kind of question than the limiting behaviour of the walk. We adopt a statistical point of view and are interested in inferring the distribution of the environment given the observation of a long trajectory of the random walk. This kind of questions has already been studied in the context of random walks in random colorings of $\mathbb{Z}$ (Benjamini and Kesten, 1996; Matzinger, 1999; Löwe and Matzinger, 2002) as well as in the context of RWRE for a characterization of the environment distribution (Adelman and Enriquez, 2004; Comets et al., 2012). Whereas Adelman and Enriquez deal with very general RWRE and present a procedure to infer the environment distribution through a system of moment equations, Comets et al. provide a maximum likelihood estimator (MLE) of the parameter of the environment distribution in the specific case of a transient ballistic one-dimensional nearest neighbour path. In the latter work, the authors establish the consistency of their estimator and provide synthetic experiments to assess its effective performances. It turns out that this estimator exhibits a much smaller variance than the one of Adelman and Enriquez. We propose to establish what the numerical investigations of Comets et al. suggested, that is, the asymptotic normality of the MLE as well as its asymptotic efficiency (namely, that it asymptotically achieves the Cramér-Rao bound).

This article is organised as follows. In Section 2.1, we introduce the framework of the one dimensional ballistic random walk in an independent and identically distributed (i.i.d.) parametric environment. In Section 2.2, we present the MLE procedure developed by Comets et al. to infer the parameter of the environment distribution. Section 2.3 recalls some already known results on an underlying branching process in a random environment related to the RWRE. Then, we state in Section 2.5 our asymptotic normality result in the wake of additional hypotheses required to prove it and listed in Section 2.4. In Section 3, we present three examples of environment distributions which are already introduced in Comets et al. (2012), and we check that the additional required assumptions of Section 2.4 are fulfilled, so that the MLE is asymptotically normal and efficient in these cases. The proof of the asymptotic normality result is presented in Section 4. We apply to the score vector sequence a central limit theorem for centered square-integrable martingales (Section 4.1) and we adapt to our context an asymptotic normality result for M-estimators (Section 4.2). To conclude this part, we provide in Section 4.3 the proof of a sufficient condition for the non-degeneracy of the Fisher information. Finally, Section 5 illustrates

our results on synthetic data by exploring empirical coverages of asymptotic confidence regions.

2 Material and results

2.1 Properties of a transient random walk in a random environment

Let us introduce a one-dimensional random walk (more precisely a nearest neighbour path) evolving in a random environment (RWRE for short) and recall its elementary properties. We start by considering the environment defined through the collection $\omega = \{\omega_x\}_{x\in\mathbb{Z}} \in (0,1)^{\mathbb{Z}}$ of i.i.d. random variables, with parametric distribution $\nu = \nu_\theta$ that depends on some unknown parameter $\theta \in \Theta$. We further assume that $\Theta \subset \mathbb{R}^d$ is a compact set. We let $\mathbb{P}^\theta = \nu_\theta^{\otimes\mathbb{Z}}$ be the law on $(0,1)^{\mathbb{Z}}$ of the environment ω and $\mathbb{E}^\theta$ be the corresponding expectation.

Now, for fixed environment ω, let $X = \{X_t\}_{t\in\mathbb{N}}$ be the Markov chain on $\mathbb{Z}$ starting at $X_0 = 0$ and with (conditional) transition probabilities

$$P_\omega(X_{t+1} = y | X_t = x) = \begin{cases} \omega_x & \text{if } y = x+1, \\ 1-\omega_x & \text{if } y = x-1, \\ 0 & \text{otherwise.} \end{cases}$$

The *quenched* distribution P_ω is the conditional measure on the path space of X given ω. Moreover, the *annealed* distribution of X is given by

$$\mathbf{P}^\theta(\cdot) = \int P_\omega(\cdot)\mathrm{d}\mathbb{P}^\theta(\omega).$$

We write E_ω and $\mathbf{E}^\theta$ for the corresponding quenched and annealed expectations, respectively. In the following, we assume that the process X is generated under the true parameter value $\theta^\star$, an interior point of the parameter space Θ, that we aim at estimating. We shorten to $\mathbf{P}^\star$ and $\mathbf{E}^\star$ (resp. $\mathbb{P}^\star$ and $\mathbb{E}^\star$) the annealed (resp. quenched) probability $\mathbf{P}^{\theta^\star}$ (resp. $\mathbb{P}^{\theta^\star}$) and corresponding expectation $\mathbf{E}^{\theta^\star}$ (resp. $\mathbb{E}^{\theta^\star}$) under parameter value $\theta^\star$.

The behaviour of the process X is related to the ratio sequence

$$\rho_x = \frac{1-\omega_x}{\omega_x}, \qquad x \in \mathbb{Z}. \tag{1}$$

We refer to Solomon (1975) for the classification of X between transient or recurrent cases according to whether $\mathbb{E}^\theta(\log \rho_0)$ is different or not from zero (the classification is also recalled in Comets et al., 2012). In our setup, we consider a transient process and without loss of generality, assume that it is transient to the right, thus corresponding to $\mathbb{E}^\theta(\log \rho_0) < 0$. The transient case may be further split into two sub-cases, called *ballistic* and *sub-ballistic* that correspond to

a linear and a sub-linear speed for the walk, respectively. More precisely, letting T_n be the first hitting time of the positive integer n,

$$T_n = \inf\{t \in \mathbb{N} : X_t = n\}, \tag{2}$$

and assuming $\mathbb{E}^\theta(\log \rho_0) < 0$ all through, we can distinguish the following cases

(a1) (Ballistic). If $\mathbb{E}^\theta(\rho_0) < 1$, then, $\mathbf{P}^\theta$-almost surely,

$$\frac{T_n}{n} \xrightarrow[n\to\infty]{} \frac{1+\mathbb{E}^\theta(\rho_0)}{1-\mathbb{E}^\theta(\rho_0)}. \tag{3}$$

(a2) (Sub-ballistic). If $\mathbb{E}^\theta(\rho_0) \geq 1$, then $T_n/n \to +\infty$, $\mathbf{P}^\theta$-almost surely when n tends to infinity.

Moreover, the fluctuations of T_n depend in nature on a parameter $\kappa \in (0,\infty]$, which is defined as the unique positive solution of

$$\mathbb{E}^\theta(\rho_0^\kappa) = 1$$

when such a number exists, and $\kappa = +\infty$ otherwise. The ballistic case corresponds to $\kappa > 1$. Under mild additional assumptions, Kesten et al. (1975) proved that

(aI) if $\kappa \geq 2$, then T_n has Gaussian fluctuations. Precisely, if c denotes the limit in (3), then $n^{-1/2}(T_n - nc)$ when $\kappa > 2$, and $(n\log n)^{-1/2}(T_n - nc)$ when $\kappa = 2$ have a non-degenerate Gaussian limit.

(aII) if $\kappa < 2$, then $n^{-1/\kappa}(T_n - d_n)$ has a non-degenerate limit distribution, which is a stable law with index κ.
The centering is $d_n = 0$ for $\kappa < 1$, $d_n = an\log n$ for $\kappa = 1$, and $d_n = an$ for $\kappa \in (1,2)$, for some positive constant a.

2.2 A consistent estimator

We briefly recall the definition of the estimator proposed in Comets et al. (2012) to infer the parameter θ, when we observe $X_{[0,T_n]} = \{X_t : t = 0,1,\ldots,T_n\}$, for some value $n \geq 1$. It is defined as the maximizer of some well-chosen criterion function, which roughly corresponds to the log-likelihood of the observations.

We start by introducing the statistics $\{L_x^n\}_{x\in\mathbb{Z}}$, defined as

$$L_x^n := \sum_{s=0}^{T_n-1} \mathbf{1}_{\{X_s=x;\ X_{s+1}=x-1\}},$$

namely L_x^n is the number of left steps of the process $X_{[0,T_n]}$ from site x. Here, $\mathbf{1}_{\{\cdot\}}$ denotes the indicator function.

Definition 2.1. *Let ϕ_θ be the function from $\mathbb{N}^2$ to $\mathbb{R}$ given by*

$$\phi_\theta(x,y) = \log \int_0^1 a^{x+1}(1-a)^y \mathrm{d}\nu_\theta(a). \tag{4}$$

The criterion function $\theta \mapsto \ell_n(\theta)$ is defined as

$$\ell_n(\theta) = \sum_{x=0}^{n-1} \phi_\theta(L_{x+1}^n, L_x^n). \tag{5}$$

We now recall the assumptions stated in Comets et al. (2012) ensuring that the maximizer of criterion ℓ_n is a consistent estimator of the unknown parameter.

Assumption I. *(Consistency conditions).*

i) (Transience to the right). For any $\theta \in \Theta, \mathbb{E}^\theta|\log \rho_0| < \infty$ and $\mathbb{E}^\theta(\log \rho_0) < 0$.

ii) (Ballistic case). For any $\theta \in \Theta$, $\mathbb{E}^\theta(\rho_0) < 1$.

iii) (Continuity). For any $(x,y) \in \mathbb{N}^2$, the map $\theta \mapsto \phi_\theta(x,y)$ is continuous on the parameter set Θ.

iv) (Identifiability). For any $(\theta,\theta') \in \Theta^2$, $\nu_\theta \neq \nu_{\theta'} \iff \theta \neq \theta'$.

v) The collection of probability measures $\{\nu_\theta : \theta \in \Theta\}$ is such that

$$\inf_{\theta\in\Theta} \mathbb{E}^\theta[\log(1-\omega_0)] > -\infty.$$

According to Assumption I point iii), the function $\theta \mapsto \ell_n(\theta)$ is continuous on the compact parameter set Θ. Thus, it achieves its maximum, and the estimator $\widehat{\theta}_n$ is defined as one maximizer of this criterion.

Definition 2.2. *An estimator $\widehat{\theta}_n$ of θ is defined as a measurable choice*

$$\widehat{\theta}_n \in \operatorname*{Argmax}_{\theta\in\Theta} \ell_n(\theta). \tag{6}$$

Note that $\widehat{\theta}_n$ is not necessarily unique. As explained in Comets et al. (2012), with a slight abuse of notation, $\widehat{\theta}_n$ may be considered a MLE. Moreover, under Assumption I, Comets et al. (2012) establish its consistency, namely its convergence in $\mathbf{P}^\star$-probability to the true parameter value $\theta^\star$.

2.3 The role of an underlying branching process

We introduce in this section an underlying branching process with immigration in random environment (BPIRE) that is naturally related to the RWRE. Indeed, it

is well-known that for an i.i.d. environment, under the annealed law $\mathbf{P}^\star$, the sequence $L_n^n, L_{n-1}^n, \dots, L_0^n$ has the same distribution as a BPIRE denoted $Z_0, \dots, Z_n$, and defined by

$$Z_0 = 0, \quad \text{and for } k = 0, \dots, n-1, \quad Z_{k+1} = \sum_{i=0}^{Z_k} \xi'_{k+1,i}, \tag{7}$$

with $\{\xi'_{k,i}\}_{k\in\mathbb{N}^*; i\in\mathbb{N}}$ independent and

$$\forall m \in \mathbb{N}, \quad P_\omega(\xi'_{k,i} = m) = (1-\omega_k)^m \omega_k,$$

(see for instance Kesten et al., 1975; Comets et al., 2012). Let us introduce through the function ϕ_θ defined by (4) the transition kernel Q_θ on $\mathbb{N}^2$ defined as

$$Q_\theta(x, y) = \binom{x+y}{x} e^{\phi_\theta(x,y)} = \binom{x+y}{x} \int_0^1 a^{x+1}(1-a)^y \mathrm{d}\nu_\theta(a). \tag{8}$$

Then for each value $\theta \in \Theta$, under annealed law $\mathbf{P}^\theta$ the BPIRE $\{Z_n\}_{n\in\mathbb{N}}$ is an irreducible positive recurrent homogeneous Markov chain with transition kernel Q_θ and a unique stationary probability distribution denoted by π_θ. Moreover, the moments of π_θ may be characterised through the distribution of the ratios $\{\rho_x\}_{x\in\mathbb{Z}}$. The following statement is a direct consequence from the proof of Theorem 4.5 in Comets et al. (2012) (see Equation (16) in this proof).

Proposition 2.3 (Theorem 4.5 in Comets et al. (2012))**.** *The invariant probability measure π_θ is positive on $\mathbb{N}$ and satisfies*

$$\forall j \geq 0, \quad \sum_{k\geq j+1} k(k-1)\dots(k-j)\pi_\theta(k) = (j+1)!\, \mathbb{E}^\theta\Big[\Big(\sum_{n\geq 1}\prod_{k=1}^n \rho_k\Big)^{j+1}\Big].$$

In particular, π_θ has a finite first moment in the ballistic case.

Note that the criterion ℓ_n satisfies the following property

$$\ell_n(\theta) \sim \sum_{k=0}^{n-1} \phi_\theta(Z_k, Z_{k+1}) \text{ under } \mathbf{P}^\star, \tag{9}$$

where $\sim$ means equality in distribution. For each value $\theta \in \Theta$, under annealed law $\mathbf{P}^\theta$ the process $\{(Z_n, Z_{n+1})\}_{n\in\mathbb{N}}$ is also an irreducible positive recurrent homogeneous Markov chain with unique stationary probability distribution denoted by $\tilde\pi_\theta$ and defined as

$$\tilde\pi_\theta(x, y) = \pi_\theta(x) Q_\theta(x, y), \quad \forall (x, y) \in \mathbb{N}^2. \tag{10}$$

For any function $g : \mathbb{N}^2 \to \mathbb{R}$ such that $\sum_{x,y} \tilde\pi_\theta(x, y)|g(x, y)| < \infty$, we denote by $\tilde\pi_\theta(g)$ the quantity

$$\tilde\pi_\theta(g) = \sum_{(x,y)\in\mathbb{N}^2} \tilde\pi_\theta(x, y) g(x, y). \tag{11}$$

We extend the notation above for any function $g = (g_1, \dots, g_d) : \mathbb{N}^2 \to \mathbb{R}^d$ such that $\tilde{\pi}_\theta(\|g\|) < \infty$, where $\|\cdot\|$ is the uniform norm, and denote by $\tilde{\pi}_\theta(g)$ the vector $(\tilde{\pi}_\theta(g_1), \dots, \tilde{\pi}_\theta(g_d))$. The following ergodic theorem is valid.

Proposition 2.4. *(Theorem 4.2 in Chapter 4 from Revuz, 1984). Under point i) in Assumption I, for any function $g : \mathbb{N}^2 \to \mathbb{R}^d$, such that $\tilde{\pi}_\theta(\|g\|) < \infty$ the following ergodic theorem holds*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} g(Z_k, Z_{k+1}) = \tilde{\pi}_\theta(g),$$

$\mathbf{P}^\star$*-almost surely and in* $\mathbb{L}^1(\mathbf{P}^\star)$.

2.4 Assumptions for asymptotic normality

Assumption I is required for the construction of a consistent estimator of the parameter θ. It mainly consists in a transient random walk with linear speed (ballistic regime) plus some regularity assumptions on the model with respect to $\theta \in \Theta$. Now, asymptotic normality result for this estimator requires additional hypotheses.

In the following, for any function g_θ depending on the parameter θ, the symbols $\dot{g}_\theta$ or $\partial_\theta g_\theta$ and $\ddot{g}_\theta$ or $\partial^2_\theta g_\theta$ denote the (column) gradient vector and Hessian matrix with respect to θ, respectively. Moreover, $Y^\intercal$ is the row vector obtained by transposing the column vector Y.

Assumption II. *(Differentiability). The collection of probability measures $\{\nu_\theta : \theta \in \Theta\}$ is such that for any $(x, y) \in \mathbb{N}^2$, the map $\theta \mapsto \phi_\theta(x, y)$ is twice continuously differentiable on Θ.*

Assumption III. *(Regularity conditions). For any $\theta \in \Theta$, there exists some $q > 1$ such that*

$$\tilde{\pi}_\theta\left(\|\dot{\phi}_\theta\|^{2q}\right) < +\infty. \tag{12}$$

For any $x \in \mathbb{N}$,

$$\sum_{y\in\mathbb{N}} \dot{Q}_\theta(x, y) = \partial_\theta \sum_{y\in\mathbb{N}} Q_\theta(x, y) = 0. \tag{13}$$

Assumption IV. *(Uniform conditions). For any $\theta \in \Theta$, there exists some neighborhood $\mathcal{V}(\theta)$ of θ such that*

$$\tilde{\pi}_\theta\Big(\sup_{\theta'\in\mathcal{V}(\theta)} \|\dot{\phi}_{\theta'}\|^2\Big) < +\infty \quad \textit{and} \quad \tilde{\pi}_\theta\Big(\sup_{\theta'\in\mathcal{V}(\theta)} \|\ddot{\phi}_{\theta'}\|^2\Big) < +\infty. \tag{14}$$

Assumptions II and III are technical and involved in the proof of a central limit theorem (CLT) for the gradient vector of the criterion ℓ_n, also called score vector

sequence. Assumption IV is also technical and involved in the proof of asymptotic normality of $\widehat{\theta}_n$ from the latter CLT. Note that Assumption III also allows us to define the matrix

$$\Sigma_\theta = \tilde{\pi}_\theta\left(\dot{\phi}_\theta \dot{\phi}_\theta^\intercal\right). \tag{15}$$

Combining the definitions (8),(10),(11) and (15) with Assumption III, we obtain the equivalent expression for Σ_θ

$$\begin{aligned}
\Sigma_\theta &= \sum_{x\in\mathbb{N}}\sum_{y\in\mathbb{N}} \pi_\theta(x) \frac{1}{Q_\theta(x,y)} \dot{Q}_\theta(x,y)\dot{Q}_\theta(x,y)^\intercal \\
&= -\sum_{x\in\mathbb{N}}\sum_{y\in\mathbb{N}} \pi_\theta(x) \left(\ddot{Q}_\theta(x,y) - \frac{1}{Q_\theta(x,y)} \dot{Q}_\theta(x,y)\dot{Q}_\theta(x,y)^\intercal\right) \\
&= -\tilde{\pi}_\theta(\ddot{\phi}_\theta).
\end{aligned} \tag{16}$$

Assumption V. *(Fisher information matrix). For any value $\theta \in \Theta$, the matrix Σ_θ is non singular.*

Assumption V states invertibility of the Fisher information matrix $\Sigma_{\theta^\star}$. This assumption is necessary to prove asymptotic normality of $\widehat{\theta}_n$ from the previously mentioned CLT on the score vector sequence.

2.5 Results

Theorem 2.5. *Under Assumptions I to III, the score vector sequence $\dot{\ell}_n(\theta^\star)/\sqrt{n}$ is asymptotically normal with mean zero and finite covariance matrix $\Sigma_{\theta^\star}$.*

Theorem 2.6. *(Asymptotic normality). Under Assumptions I to V, for any choice of $\widehat{\theta}_n$ satisfying (6), the sequence $\{\sqrt{n}(\widehat{\theta}_n - \theta^\star)\}_{n\in\mathbb{N}}$ converges in $\mathbf{P}^\star$-distribution to a centered Gaussian random vector with covariance matrix $\Sigma_{\theta^\star}^{-1}$.*

Note that the limiting covariance matrix of $\sqrt{n}\widehat{\theta}_n$ is exactly the inverse Fisher information matrix of the model. As such, our estimator is efficient. Moreover, the previous theorem may be used to build asymptotic confidence regions for θ, as illustrated in Section 5. In this section, we also explain how to estimate the Fisher information matrix $\Sigma_{\theta^\star}$. Indeed, $\Sigma_{\theta^\star}$ is defined via the invariant distribution $\tilde{\pi}_{\theta^\star}$ which possesses no analytical expression. To bypass the problem, we rely on the observed Fisher information matrix as an estimator of $\Sigma_{\theta^\star}$.

Remark 2.7. *We observe that the fluctuations of the estimator $\widehat{\theta}_n$ are unrelated to those of T_n or those of X_t, see (aI)-(aII). Though there is a change of limit law from Gaussian to stable as $\mathbb{E}^\theta(\rho_0^2)$ decreases from larger to smaller than 1, the MLE remains asymptotically normal in the full ballistic region (no extra assumption is required in Example I introduced in Section 3). We illustrate this point by considering a naive estimator at the end of Subsection 3.1.*

We conclude this section by providing a sufficient condition for Assumption V to be valid, namely ensuring that Σ_θ is positive definite.

Proposition 2.8. *For the covariance matrix Σ_θ to be positive definite, it is sufficient that the linear span in $\mathbb{R}^d$ of the gradient vectors $\dot{\phi}_\theta(x,y)$, with $(x,y)\in\mathbb{N}^2$ is equal to the full space, or equivalently, that*

$$\mathrm{Vect}\Big\{\partial_\theta \mathbb{E}^\theta(\omega_0^{x+1}(1-\omega_0)^y) : (x,y)\in\mathbb{N}^2\Big\} = \mathbb{R}^d.$$

Section 4 is devoted to the proof of Theorem 2.6 where Subsections 4.1 and 4.3 are concerned with the proofs of Theorem 2.5 and Proposition 2.8, respectively.

3 Examples

3.1 Environment with finite and known support

Example I. *Fix $a_1 < a_2 \in (0,1)$ and let $\nu_p = p\delta_{a_1} + (1-p)\delta_{a_2}$, where δ_a is the Dirac mass located at value a. Here, the unknown parameter is the proportion $p\in\Theta\subset[0,1]$ (namely $\theta = p$). We suppose that a_1, a_2 and Θ are such that points $i)$ and $ii)$ in Assumption I are satisfied.*

This example is easily generalized to ν having $m\geq 2$ support points namely $\nu_\theta = \sum_{i=1}^m p_i a_i$, where $a_1,\ldots,a_m$ are distinct, fixed and known in $(0,1)$, we let $p_m = 1-\sum_{i=1}^{m-1} p_i$ and the parameter is now $\theta = (p_1,\ldots,p_{m-1})$.

In the framework of Example I, we have

$$\phi_p(x,y) = \log[pa_1^{x+1}(1-a_1)^y + (1-p)a_2^{x+1}(1-a_2)^y], \tag{17}$$

and

$$\ell_n(p) := \ell_n(\theta) = \sum_{x=0}^{n-1} \log\Big[pa_1^{L_{x+1}^n+1}(1-a_1)^{L_x^n} + (1-p)a_2^{L_{x+1}^n+1}(1-a_2)^{L_x^n}\Big]. \tag{18}$$

Comets et al. (2012) proved that $\widehat{p}_n = \mathrm{Argmax}_{p\in\Theta}\,\ell_n(p)$ converges in $\mathbf{P}^\star$-probability to $p^\star$. There is no analytical expression for the value of $\widehat{p}_n$. Nonetheless, this estimator may be easily computed by numerical methods. We now establish that the assumptions needed for asymptotic normality are also satisfied in this case, under the only additional assumption that $\Theta\subset(0,1)$.

Proposition 3.1. *In the framework of Example I, assuming moreover that $\Theta\subset(0,1)$, Assumptions II to IV are satisfied.*

Proof. The function $p \mapsto \phi_p(x,y)$ given by (17) is twice continuously differentiable for any (x,y). The derivatives are given by

$$\dot{\phi}_p(x,y) = \mathrm{e}^{-\phi_p(x,y)}[a_1^{x+1}(1-a_1)^y - a_2^{x+1}(1-a_2)^y],$$
$$\ddot{\phi}_p(x,y) = -\dot{\phi}_p(x,y)^2.$$

Since $\exp[\phi_p(x,y)] \geq p a_1^{x+1}(1-a_1)^y$ and $\exp[\phi_p(x,y)] \geq (1-p)a_2^{x+1}(1-a_2)^y$, we obtain the bounds

$$|\dot{\phi}_p(x,y)| \leq \frac{1}{p} + \frac{1}{1-p}.$$

Now, under the additional assumption that $\Theta \subset (0,1)$, there exists some $A \in (0,1)$ such that $\Theta \subset [A, 1-A]$ and then

$$\sup_{(x,y)\in\mathbb{N}^2} |\dot{\phi}_p(x,y)| \leq \frac{2}{A} \quad \text{and} \quad \sup_{(x,y)\in\mathbb{N}^2} |\ddot{\phi}_p(x,y)| \leq \frac{4}{A^2}, \tag{19}$$

which yields that (12) and (14) are satisfied.

Now, noting that

$$\dot{Q}_\theta(x,y) = \binom{x+y}{x}[a_1^{x+1}(1-a_1)^y - a_2^{x+1}(1-a_2)^y],$$

and that

$$\sum_{y=0}^{\infty} \binom{x+y}{x} a^{x+1}(1-a)^y = 1, \quad \forall x \in \mathbb{N},\ \forall a \in (0,1), \tag{20}$$

yields (13). □

Proposition 3.2. *In the framework of Example I, the covariance matrix Σ_θ is positive definite, namely Assumption V is satisfied.*

Proof of Proposition 3.2. We have

$$\mathbb{E}^p(\omega_0) = p(a_1 - a_2) + a_2,$$

with derivative $a_1 - a_2 \neq 0$, which achieves the proof thanks to Proposition 2.8. □

Thanks to Theorem 2.6 and Propositions 3.1 and 3.2, the sequence $\{\sqrt{n}(\widehat{p}_n - p^\star)\}$ converges in $\mathbf{P}^\star$-distribution to a non degenerate centered Gaussian random variable, with variance

$$\Sigma_{p^\star}^{-1} = \left\{ \sum_{(x,y)\in\mathbb{N}^2} \pi_{p^\star}(x) \binom{x+y}{x} \frac{[a_1^{x+1}(1-a_1)^y - a_2^{x+1}(1-a_2)^y]^2}{p^\star a_1^{x+1}(1-a_1)^y + (1-p^\star)a_2^{x+1}(1-a_2)^y} \right\}^{-1}.$$

Remark 3.3. *(Temkin model, cf. Hughes (1996)) With $a \in (1/2, 1)$ known and $\theta = p \in (0,1)$ unknown, we consider $\nu_\theta = p\delta_a + (1-p)\delta_{1-a}$. This is a particular case of Example I. It is easy to see that transience to the right and ballistic regime, respectively, are equivalent to*

$$p > 1/2, \qquad p > a,$$

and that in the ballistic case, the limit $c = c(p)$ in (3) is given by

$$c(p) = \frac{a + p - 2ap}{(2a-1)(p-a)}.$$

We construct a new estimator $\tilde{p}_n$ of p solving the relation $c(\tilde{p}_n) = T_n/n$, namely

$$\tilde{p}_n = \frac{a}{2a-1} \times \frac{(2a-1)T_n + n}{T_n + n}.$$

This new estimator is consistent in the full ballistic region. However, for all $a > 1/2$ and $p > a$ but close to it, we have $\kappa \in (1,2)$, the fluctuations of T_n are of order $n^{1/\kappa}$, and those of $\tilde{p}_n$ are of the same order. This new estimator is much more spread out than the MLE $\widehat{p}_n$.

3.2 Environment with two unknown support points

Example II. *We let $\nu_\theta = p\delta_{a_1} + (1-p)\delta_{a_2}$ and now the unknown parameter is $\theta = (p, a_1, a_2) \in \Theta$, where Θ is a compact subset of*

$$(0,1) \times \{(a_1, a_2) \in (0,1)^2 \, : \, a_1 < a_2\}.$$

We suppose that Θ is such that points $i)$ and $ii)$ in Assumption I are satisfied.

The function ϕ_θ and the criterion $\ell_n(\cdot)$ are given by (17) and (18), respectively. Comets et al. (2012) have established that the estimator $\widehat{\theta}_n$ is well-defined and consistent in probability. Once again, there is no analytical expression for the value of $\widehat{\theta}_n$. Nonetheless, this estimator may also be easily computed by numerical methods. We now establish that the assumptions needed for asymptotic normality are also satisfied in this case, under a mild additional moment assumption.

Proposition 3.4. *In the framework of Example II, assuming moreover that $\mathbb{E}^\theta(\rho_0^3) < 1$, Assumptions II to IV are satisfied.*

Proof. In the proof of Proposition 3.1, we have already controled the derivative of $\theta \mapsto \phi_\theta(x,y)$ with respect to p. Hence, it is now sufficient to control its derivatives with respect to a_1 and a_2 to achieve the proof of (12) and (14). We have

$$\partial_{a_1}\phi_\theta(x,y) = \mathrm{e}^{-\phi_\theta(x,y)} p a_1^x (1-a_1)^{y-1}[(x+1)(1-a_1) - ya_1],$$
$$\partial_{a_2}\phi_\theta(x,y) = \mathrm{e}^{-\phi_\theta(x,y)} (1-p) a_2^x (1-a_2)^{y-1}[(x+1)(1-a_2) - ya_2].$$

Since

$$\mathrm{e}^{-\phi_\theta(x,y)} p a_1^x (1-a_1)^{y-1} \le \frac{1}{a_1(1-a_1)},$$

and

$$\mathrm{e}^{-\phi_\theta(x,y)} (1-p) a_2^x (1-a_2)^{y-1} \le \frac{1}{a_2(1-a_2)},$$

we can see that there exists a constant B such that

$$|\partial_{a_j} \phi_\theta(x,y)| \le \left| \frac{x+1}{a_j} - \frac{y}{1-a_j} \right| \le B(x+1+y), \quad \text{for } j = 1,2. \tag{21}$$

Now, we prove that (12) is satisfied with $q = 3/2$. From (21), it is sufficient to check that

$$\sum_{k\in\mathbb{N}} k^3 \pi_\theta(k) < \infty,$$

which is equivalent to

$$\sum_{k\ge3} k(k-1)(k-2)\pi_\theta(k) = 6\mathbb{E}^\theta \Big[\Big(\sum_{n\ge1} \prod_{k=1}^n \rho_k \Big)^3 \Big] < \infty,$$

where the last equality comes from Proposition 2.3. From Minkowski's inequality, we have

$$\mathbb{E}^\theta \Big[\Big(\sum_{n\ge1} \prod_{k=1}^n \rho_k \Big)^3 \Big] \le \Big\{ \sum_{n\ge1} \Big[\mathbb{E}^\theta \Big(\prod_{k=1}^n \rho_k^3 \Big) \Big]^{1/3} \Big\}^3 = \Big\{ \sum_{n\ge1} [\mathbb{E}^\theta(\rho_0^3)]^{n/3} \Big\}^3,$$

where the right-hand side term is finite according to the additional assumption that $\mathbb{E}^\theta(\rho_0^3) < 1$. Since the bound in (21) does not depend on θ and π_θ possesses a finite third moment, the first part of condition (14) on the gradient vector is also satisfied.

Now, we turn to (13). Noting that

$$\partial_{a_1} Q_\theta(x,y) = \binom{x+y}{x} p a_1^x (1-a_1)^{y-1} [(x+1)(1-a_1) - y a_1],$$

$$\partial_{a_2} Q_\theta(x,y) = \binom{x+y}{x} (1-p) a_2^x (1-a_2)^{y-1} [(x+1)(1-a_2) - y a_2],$$

that

$$\sum_{y=0}^\infty y \binom{x+y}{x} a^{x+1} (1-a)^y = (x+1)\frac{1-a}{a}, \quad \forall x \in \mathbb{N},\ \forall a \in (0,1), \tag{22}$$

and using (20) yields (13).

The second order derivatives of ϕ_θ are given by

$$\begin{aligned}
\partial_p^2 \phi_\theta(x,y) &= -[\partial_p \phi_\theta(x,y)]^2, \\
\partial_p \partial_{a_1} \phi_\theta(x,y) &= [\partial_{a_1}\phi_\theta(x,y)] \times \left(\frac{1}{p} - \partial_p \phi_\theta(x,y)\right), \\
\partial_{a_1}\partial_{a_2}\phi_\theta(x,y) &= -[\partial_{a_1}\phi_\theta(x,y)] \times [\partial_{a_2}\phi_\theta(x,y)], \\
\partial_{a_1}^2 \phi_\theta(x,y) &= [\partial_{a_1}\phi_\theta(x,y)] \times \Big[-\partial_{a_1}\phi_\theta(x,y) + \frac{x}{a_1} - \frac{y-1}{1-a_1} \\
&\qquad - \frac{x+1+y}{(x+1)(1-a_1)-ya_1} \Big],
\end{aligned}$$

and similar formulas for a_2 instead of a_1. The second part of (14) on the Hessian matrix thus follows from the previous expressions combined with (19), (21) and the existence of a second order moment for π_θ. □

Proposition 3.5. *In the framework of Example II, the covariance matrix Σ_θ is positive definite, namely Assumption V is satisfied.*

Proof of Proposition 3.5. We have

$$\mathbb{E}^\theta[\omega_0^{x+1}(1-\omega_0)^x] = p a_1^{x+1}(1-a_1)^x + (1-p)a_2^{x+1}(1-a_2)^x.$$

The determinant of $\left(\partial_\theta \mathbb{E}^\theta[\omega_0^{k+1}(1-\omega_0)^k]\right)_{k=0,1,x}$ is given by

$$\begin{vmatrix}
a_1-a_2 & a_1^2(1-a_1)-a_2^2(1-a_2) & a_1^{x+1}(1-a_1)^x - a_2^{x+1}(1-a_2)^x \\
p & pa_1(2-3a_1) & pa_1^x(1-a_1)^{x-1}[x(1-2a_1)+1-a_1] \\
1-p & (1-p)a_2(2-3a_2) & (1-p)a_2^x(1-a_2)^{x-1}[x(1-2a_2)+1-a_2]
\end{vmatrix}$$

and we denote it by Det. As we have $a_1 \neq a_2$ and $p \in (0,1)$, we show that this determinant is non zero for large x. This will complete the proof, thanks to Proposition 2.8.

We first consider the case of $a_1(1-a_1) \neq a_2(1-a_2)$, *i.e.* of $a_2 \neq 1-a_1$ since $a_1 < a_2$. Without loss of generality, we assume $a_1(1-a_1) < a_2(1-a_2)$. In this case, the leading terms as $x \to \infty$ in Det are

$$\begin{aligned}
\mathrm{Det} = p(1-p)a_2^x(1-a_2)^{x-1} \times \\
&\begin{vmatrix}
a_1-a_2 & a_1^2(1-a_1)-a_2^2(1-a_2) & -a_2(1-a_2) \\
1 & a_1(2-3a_1) & 0 \\
1 & a_2(2-3a_2) & [x(1-2a_2)+1-a_2]
\end{vmatrix} \\
&+ o(a_2^x(1-a_2)^x).
\end{aligned}$$

The sign of Det is determined by that of the above, new determinant. By transience of the walk to $+\infty$, it holds $a_2 > 1/2$, and we see in this new determinant

that the coefficient of x, namely $(a_2 - a_1)^2(1-2a_2)(1-2a_1-a_2)$, and the constant term $a_2(a_2-1)(a_2-a_1)[2-3a_1-3a_2+(a_2-a_1)(1-2a_1-a_2)]$ do not vanish simultaneously. Therefore, $\mathrm{Det} \neq 0$ for large x.

In the case $a_2 = 1 - a_1$, with some algebra we find

$$\mathrm{Det} = p(1-p)a_1^x(1-a_1)^x(a_2-a_1)^3(1-2a_2)a_1a_2x + \mathcal{O}(a_1^x(1-a_1)^x),$$

with a nonzero leading term. Hence $\mathrm{Det} \neq 0$ for large x. $\square$

Thanks to Theorem 2.6 and Propositions 3.4 and 3.5, under the additional assumption that $\mathbb{E}^\theta(\rho_0^3) < 1$, the sequence $\{\sqrt{n}(\widehat{\theta}_n - \theta^\star)\}$ converges in $\mathbf{P}^\star$-distribution to a non degenerate centered Gaussian random vector.

3.3 Environment with Beta distribution

Example III. *We let ν be a Beta distribution with parameters (α, β), namely*

$$\mathrm{d}\nu(a) = \frac{1}{\mathrm{B}(\alpha,\beta)} a^{\alpha-1}(1-a)^{\beta-1}\mathrm{d}a, \quad \mathrm{B}(\alpha,\beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}\mathrm{d}t.$$

Here, the unknown parameter is $\theta = (\alpha, \beta) \in \Theta$ where Θ is a compact subset of

$$\{(\alpha,\beta) \in (0,+\infty)^2 : \alpha > \beta + 1\}.$$

As $\mathbb{E}^\theta(\rho_0) = \beta/(\alpha-1)$, the constraint $\alpha > \beta + 1$ ensures that points $i)$ and $ii)$ in Assumption I are satisfied.

In the framework of Example III, we have

$$\phi_\theta(x,y) = \log \frac{\mathrm{B}(x+1+\alpha, y+\beta)}{\mathrm{B}(\alpha,\beta)} \tag{23}$$

and

$$\begin{aligned}\ell_n(\theta) &= -n\log \mathrm{B}(\alpha,\beta) + \sum_{x=0}^{n-1} \log \mathrm{B}(L_{x+1}^n + \alpha + 1, L_x^n + \beta) \\ &= \sum_{x=0}^{n-1} \log \frac{(L_{x+1}^n+\alpha)(L_{x+1}^n+\alpha-1)\dots\alpha \times (L_x^n+\beta-1)(L_x^n+\beta-2)\dots\beta}{(L_{x+1}^n+L_x^n+\alpha+\beta-1)(L_{x+1}^n+L_x^n+\alpha+\beta-2)\dots(\alpha+\beta)}.\end{aligned}$$

In this case, Comets et al. (2012) proved that $\widehat{\theta}_n$ is well-defined and consistent in probability. We now establish that the assumptions needed for asymptotic normality are also satisfied in this case.

Proposition 3.6. *In the framework of Example III, Assumptions II to IV are satisfied.*

Proof of Proposition 3.6. Relying on classical identities on the Beta function, it may be seen after some computations that

$$\phi_\theta(x,y) = \sum_{k=0}^{x} \log(k+\alpha) + \sum_{k=0}^{y-1} \log(k+\beta) - \sum_{k=0}^{x+y} \log(k+\alpha+\beta).$$

As a consequence, we obtain

$$\begin{aligned} \partial_\alpha \phi_\theta(x,y) &= \sum_{k=0}^{x} \frac{1}{k+\alpha} - \sum_{k=0}^{x+y} \frac{1}{k+\alpha+\beta} \\ &= \sum_{k=0}^{x} \frac{\beta}{(k+\alpha)(k+\alpha+\beta)} - \sum_{k=1}^{y} \frac{1}{k+x+\alpha+\beta}. \end{aligned} \tag{24}$$

The fact that Θ is a compact set included in $(0,+\infty)^2$ yields the existence of a constant A independent of θ, x and y such that both

$$\sum_{k=0}^{x} \frac{\beta}{(k+\alpha)(k+\alpha+\beta)} \le \sum_{k=0}^{+\infty} \frac{\beta}{(k+\alpha)(k+\alpha+\beta)} \le A,$$

and

$$\sum_{k=1}^{y} \frac{1}{k+x+\alpha+\beta} \le \sum_{k=1}^{y} \frac{1}{k+\alpha+\beta} \le A \log(1+y).$$

The same holds for $\partial_\beta \phi_\theta(x,y)$. Hence, we have

$$|\partial_\alpha \phi_\theta(x,y)| \le A' \log(1+y) \quad \text{and} \quad |\partial_\beta \phi_\theta(x,y)| \le A' \log(1+x), \tag{25}$$

for some positive constant A'. Since there exists a constant B such that for any integer x

$$\log(1+x) \le B \sqrt[4]{x},$$

we deduce from (25) that there exists $C > 0$ such that

$$|\partial_\alpha \phi_\theta(x,y)|^{2q} \le Cy \quad \text{and} \quad |\partial_\beta \phi_\theta(x,y)|^{2q} \le Cx, \tag{26}$$

where $q = 2$. From Proposition 2.3, we know that π_θ possesses a finite first moment, and together with (26), this is sufficient for (12) to be satisfied. Since the bound in (26) does not depend on θ, the first part of condition (14) on the gradient vector is also satisfied.

Now, we prove that it is possible to exchange the order of derivation and summation to get (13). To do so, we prove that

$$\sum_{y} \|\dot{Q}_\theta(x,y)\| < \infty, \tag{27}$$

for any integer x. Define $\theta_0 = (\alpha_0, \beta_0)$ with

$$\alpha_0 = \inf(\mathrm{proj}_1(\Theta)) \quad \text{and} \quad \beta_0 = \inf(\mathrm{proj}_2(\Theta)),$$

where $\text{proj}_i, i = 1,2$ are the two projectors on the coordinates. Note that θ_0 does not necessarily belong to Θ. However, it still belongs to the ballistic region $\{\alpha > \beta + 1\}$. For any $a \in (0,1)$ and any integers x and y, we have

$$a^{x+1+\alpha-1}(1-a)^{y+\beta-1} \le a^{x+1+\alpha_0-1}(1-a)^{y+\beta_0-1},$$

which yields

$$\mathrm{B}(x+1+\alpha, y+\beta) \le \mathrm{B}(x+1+\alpha_0, y+\beta_0),$$

as well as

$$Q_\theta(x,y) \le \frac{\mathrm{B}(\alpha_0,\beta_0)}{\mathrm{B}(\alpha,\beta)} Q_{\theta_0}(x,y).$$

Using the fact that the beta function is continuous on the compact set Θ yields the existence of a constant C such that

$$Q_\theta(x,y) \le C Q_{\theta_0}(x,y),$$

for any integers x and y. Now recall that $\dot{Q}_\theta(x,y) = Q_\theta(x,y)\dot{\phi}_\theta(x,y)$. Hence, using the last inequality and (26), it is sufficient to prove that

$$\sum_y y Q_{\theta_0}(x,y) < \infty, \tag{28}$$

to get (27). We have

$$\sum_x \Big(\sum_y y Q_{\theta_0}(x,y)\Big)\pi_{\theta_0}(x) = \sum_y y\pi_{\theta_0}(y) < \infty,$$

where the last inequality comes from the fact that θ_0 lies in the ballistic region and thus π_{θ_0} possesses a finite first moment. Hence, (28) is satisfied for any integer x which proves that (27) is satisfied.

The second order derivatives of ϕ_θ are given by

$$\partial_\alpha^2 \phi_\theta(x,y) = -\sum_{k=0}^{x} \frac{1}{(k+\alpha)^2} + \sum_{k=0}^{x+y} \frac{1}{(k+\alpha+\beta)^2},$$
$$\partial_\alpha\partial_\beta \phi_\theta(x,y) = \sum_{k=0}^{x+y} \frac{1}{(k+\alpha+\beta)^2},$$

and similar formulas for β instead of α. Thus, the second part of condition (14) for the Hessian matrix follows by arguments similar to those establishing the first part of (14) for the gradient vector. □

Proposition 3.7. *In the framework of Example III, the covariance matrix Σ_θ is positive definite, namely Assumption V is satisfied.*

Proof of Proposition 3.7. One easily checks that

$$\dot{\phi}_\theta(x,x) = \begin{pmatrix} \frac{1}{\alpha+x} + \frac{1}{\alpha+x-1} + \cdots + \frac{1}{\alpha} - \frac{1}{\alpha+\beta+2x} - \frac{1}{\alpha+\beta+2x-1} - \cdots - \frac{1}{\alpha+\beta} \\ \frac{1}{\beta+x-1} + \frac{1}{\beta+x-2} + \cdots + \frac{1}{\beta} - \frac{1}{\alpha+\beta+2x} - \frac{1}{\alpha+\beta+2x-1} - \cdots - \frac{1}{\alpha+\beta} \end{pmatrix}.$$

Hence, $\dot{\phi}_\theta(0,0)$ is collinear to $(\beta,-\alpha)^\intercal$ and $\dot{\phi}_\theta(x,x) \to (-\log 2, -\log 2)^\intercal$ as $x \to \infty$. This shows that $\dot{\phi}_\theta(x,x), x \in \mathbb{N}$, spans the whole space, and Proposition 2.8 applies. □

Thanks to Theorem 2.6 and Propositions 3.6 and 3.7, the sequence $\{\sqrt{n}(\widehat{\theta}_n - \theta^\star)\}$ converges in $\mathbf{P}^\star$-distribution to a non degenerate centered Gaussian random vector.

4 Asymptotic normality

We now establish the asymptotic normality of $\widehat{\theta}_n$ stated in Theorem 2.6. The most important step lies in establishing Theorem 2.5 that states a CLT for the gradient vector of the criterion ℓ_n (see Section 4.1). As a consequence, we obtain the asymptotic normality of $\widehat{\theta}_n$, following the proof of Theorem 5.23 in van der Vaart (1998). This latter reference deals with i.i.d. observations only, but may be easily generalized to our context as explained in Section 4.2. Finally Section 4.3 establishes the proof of Proposition 2.8 stating a condition under which the Fisher information matrix is non singular.

4.1 A central limit theorem for the gradient of the criterion

In this section, we prove Theorem 2.5, that is, the existence of a CLT for the score vector sequence $\dot{\ell}_n(\theta^\star)$. Note that according to (9), we have

$$\frac{1}{\sqrt{n}} \dot{\ell}_n(\theta^\star) \sim \frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} \dot{\phi}_{\theta^\star}(Z_k, Z_{k+1}), \tag{29}$$

where $\{Z_k\}_{0\le k\le n}$ is the Markov chain introduced in Section 2.3. First, note that under Assumption III this quantity is integrable and centered with respect to $\mathbf{P}^\star$. Indeed, recall that $\dot{\phi}_\theta(x,y) = \dot{Q}_\theta(x,y)/Q_\theta(x,y)$, thus we can write for all $x \in \mathbb{N}$,

$$\begin{aligned} \mathbf{E}^\star(\dot{\phi}_{\theta^\star}(Z_k, Z_{k+1}) | Z_k = x) &= \sum_{y\in\mathbb{N}} \frac{\dot{Q}_{\theta^\star}(x,y)}{Q_{\theta^\star}(x,y)} Q_{\theta^\star}(x,y) = \partial_\theta \Big(\sum_{y\in\mathbb{N}} Q_\theta(x,y) \Big) \Big|_{\theta=\theta^\star} \\ &= \partial_\theta(1)\Big|_{\theta=\theta^\star} = 0, \end{aligned} \tag{30}$$

where we have used (13) to interchange sum and derivative. Then,

$$\mathbf{E}^\star(\dot{\phi}_{\theta^\star}(Z_k, Z_{k+1})) = 0.$$

Now, we rely on a CLT for centered square-integrable martingales, see Theorem 3.2 in Hall and Heyde (1980). We introduce the quantities

$$\forall 1 \le k \le n, \quad U_{n,k} = \frac{1}{\sqrt{n}} \dot{\phi}_{\theta^\star}(Z_{k-1}, Z_k) \quad \text{and} \quad S_{n,k} = \sum_{j=1}^{k} U_{n,j},$$

as well as the natural filtration $\mathscr{F}_{n,k} = \mathscr{F}_k := \sigma(Z_j, j \le k)$. According to (30), $\{S_{n,k}, 1 \le k \le n, n \ge 1\}$ is a martingale array with differences $U_{n,k}$. It is also centered and square integrable from Assumption III. Thus according to Theorem 3.2 in Hall and Heyde (1980), as soon as we have

$$\max_{1 \le i \le n} \|U_{n,i}\| \xrightarrow[n \to +\infty]{} 0 \text{ in } \mathbf{P}^\star\text{-probability}, \tag{31}$$

$$\sum_{i=1}^{n} U_{n,i} U_{n,i}^\intercal \xrightarrow[n \to +\infty]{} \Sigma_{\theta^\star} \text{ in } \mathbf{P}^\star\text{-probability}, \tag{32}$$

$$\text{and } \{\mathbf{E}^\star(\max_{1 \le i \le n} U_{n,i} U_{n,i}^\intercal)\}_{n \in \mathbb{N}} \text{ is a bounded sequence}, \tag{33}$$

with $\Sigma_{\theta^\star}$ a deterministic and finite covariance matrix, then the sum $S_{n,n}$ converges in distribution to a centered Gaussian random variable with covariance matrix $\Sigma_{\theta^\star}$, which proves Theorem 2.5. Now, the convergence (32) is a direct consequence of the ergodic theorem stated in Proposition 2.4. Moreover the limit $\Sigma_{\theta^\star}$ is given by (15) and is finite according to Assumption III. Note that more generally, the ergodic theorem (Proposition 2.4) combined with Assumption III implies the convergence of $\{\sum_{1 \le i \le n} \|U_{n,i}\|^2\}_n$ to a finite deterministic limit, $\mathbf{P}^\star$-almost surely and in $\mathbb{L}_1(\mathbf{P}^\star)$. Thus, condition (33) follows from this $\mathbb{L}_1(\mathbf{P}^\star)$-convergence, combined with the bound

$$\|\mathbf{E}^\star(\max_{1 \le i \le n} U_{n,i} U_{n,i}^\intercal)\| \le \sum_{i=1}^{n} \mathbf{E}^\star(\|U_{n,i}\|^2).$$

Finally, condition (31) is obtained by writing that for any $\varepsilon > 0$ and any $q > 1$, we have

$$\begin{aligned}
\mathbf{P}^\star(\max_{1 \le i \le n} \|U_{n,i}\| \ge \varepsilon) &= \mathbf{P}^\star(\max_{1 \le i \le n} \|\dot{\phi}_{\theta^\star}(Z_{i-1}, Z_i)\| \ge \varepsilon\sqrt{n}) \\
&\le \frac{1}{n^q \varepsilon^{2q}} \mathbf{E}^\star(\max_{1 \le i \le n} \|\dot{\phi}_{\theta^\star}(Z_{i-1}, Z_i)\|^{2q}) \\
&\le \frac{1}{n^q \varepsilon^{2q}} \sum_{i=1}^{n} \mathbf{E}^\star(\|\dot{\phi}_{\theta^\star}(Z_{i-1}, Z_i)\|^{2q}),
\end{aligned}$$

where the first inequality is Markov's inequality. By using again Assumption III and the ergodic theorem (Proposition 2.4), the right-hand side of this inequality converges to zero whenever $q > 1$. This achieves the proof.

4.2 Proof of asymptotic normality

We follow the proof of asymptotic normality result for M-estimators stated in Theorem 5.23 in van der Vaart (1998) in a i.i.d. context. Indeed, our estimator $\widehat{\theta}_n$ maximizes the function $\theta \mapsto \ell_n(\theta) = \sum_{x=0}^{n-1} \phi_\theta(L_{x+1}^n, L_x^n)$ and converges in $\mathbf{P}^\star$-probability to $\theta^\star$. Moreover, it is shown in Comets et al. (2012) that the normalised criterion ℓ_n/n satisfies

$$\frac{1}{n}\ell_n(\theta) \xrightarrow[n\to+\infty]{} \ell(\theta) := \tilde{\pi}_{\theta^\star}(\phi_\theta),$$

in $\mathbf{P}^\star$-probability and the limiting function ℓ has a unique maximum at $\theta^\star$ (see Theorem 4.1 and Section 4.4 in Comets et al., 2012). Under Assumption III, we obtain the following

1) The function $\theta \mapsto \phi_\theta(x,y)$ is differentiable at $\theta^\star$ for all (x,y) and there exists some positive function $\dot{\phi} : \mathbb{N}^2 \to \mathbb{R}^d$ such that

$$\tilde{\pi}_{\theta^\star}(\dot{\phi}\dot{\phi}^\intercal) < +\infty$$

and for any θ_1, θ_2 in a neighborhood of $\theta^\star$, for any $(x,y) \in \mathbb{N}^2$, we have

$$|\phi_{\theta_1}(x,y) - \phi_{\theta_2}(x,y)| \le \|\dot{\phi}(x,y)\| \cdot \|\theta_1 - \theta_2\|.$$

2) The map $\theta \mapsto \ell(\theta)$ admits a second order Taylor expansion at its maximum $\theta^\star$.

If we moreover assume that the Fisher information matrix $\Sigma_{\theta^\star} = -\tilde{\pi}_{\theta^\star}(\ddot{\phi}_{\theta^\star})$ is non singular, then we have

$$\sqrt{n}(\widehat{\theta}_n - \theta^\star) = \Sigma_{\theta^\star}^{-1} \frac{1}{\sqrt{n}} \sum_{x=0}^{n-1} \dot{\phi}_{\theta^\star}(L_{x+1}^n, L_x^n) + o_P(1), \tag{34}$$

where $o_P(1)$ is a remainder term that converges in $\mathbf{P}^\star$-probability to 0. The proof of the latter fact is a simple rewriting of the proof of Theorem 5.23 in van der Vaart (1998) and is therefore omitted. The main point is that the usual empirical process $\mathbb{G}_n$ appearing in the original proof should be replaced here by its counterpart in our framework, namely the operator

$$\tilde{\mathbb{G}}_n(\phi) := \frac{1}{\sqrt{n}} \sum_{x=0}^{n-1} \Big\{ \phi(L_{x+1}^n, L_x^n) - \tilde{\pi}_\theta(\phi) \Big\},$$

for any $\phi : \mathbb{N}^2 \mapsto \mathbb{R}$ or $\mathbb{R}^d$ such that $\tilde{\pi}_\theta(\|\phi\|) < +\infty$. Combining the equality in distribution between $L_n^n, L_{n-1}^n, \ldots, L_0^n$ and the positive recurrent Markov chain $\{Z_k\}_{0\le k\le n}$ with the ergodic theorem (Proposition 2.4) applied to this latter Markov chain, the operator $\tilde{\mathbb{G}}_n$ satisfies

$$\frac{1}{\sqrt{n}}\tilde{\mathbb{G}}_n(\phi) = o_P(1),$$

which is the main ingredient of the proof.

Finally, combining (34) with Theorem 2.5, we obtain the convergence in $\mathbf{P}^\star$- distribution of $\{\sqrt{n}(\widehat{\theta}_n - \theta^\star)\}$ to a centered Gaussian random vector with covariance matrix $\Sigma_{\theta^\star}^{-1}\Sigma_{\theta^\star}\Sigma_{\theta^\star}^{-1} = \Sigma_{\theta^\star}^{-1}$.

4.3 Non degeneracy of the Fisher information

We now turn to the proof of Proposition 2.8. Let us consider a deterministic vector $u \in \mathbb{R}^d$. We have

$$u^\intercal \Sigma_\theta u = \tilde{\pi}_\theta(\| u^\intercal \dot{\phi}_\theta \|^2).$$

We recall that according to Proposition 2.3, the invariant probability measure π_θ is positive as well as $\tilde{\pi}_\theta$. As a consequence, the quantity $u^\intercal \Sigma_\theta u$ is non negative and equals zero if and only if

$$\forall x, y \in \mathbb{N}, \quad u^\intercal \dot{\phi}_\theta(x, y) = 0.$$

Let us assume that the linear span in $\mathbb{R}^d$ of the gradient vectors $\dot{\phi}_\theta(x, y), (x, y) \in \mathbb{N}^2$ is equal to the full space, or equivalently, that

$$\mathrm{Vect}\Big\{\partial_\theta \mathbb{E}^\theta(\omega_0^{x+1}(1-\omega_0)^y) : (x, y) \in \mathbb{N}^2\Big\} = \mathbb{R}^d.$$

Then, the equality $u^\intercal \dot{\phi}_\theta(x, y) = 0$ for any $(x, y) \in \mathbb{N}^2$ implies $u = 0$. This concludes the proof.

5 Numerical performances

In Comets et al. (2012), the authors have investigated the numerical performances of the MLE and obtained that this estimator has better performances than the one proposed by Adelman and Enriquez (2004), being less spread out than the latter. In this section, we explore the possibility to construct confidence regions for the parameter θ, relying on the asymptotic normality result obtained in Theorem 2.6. Indeed, the limiting covariance $\Sigma_{\theta^\star}^{-1}$ may be approximated by the *observed Fisher information matrix*

$$\hat{\Sigma}_n = -\frac{1}{n}\sum_{x=0}^{n-1} \ddot{\phi}_{\widehat{\theta}_n}(L_{x+1}^n, L_x^n). \tag{35}$$

The consistency of $\widehat{\theta}_n$ combined with Proposition 2.4, Theorem 2.6 and Slutsky's Lemma first gives the convergence of $\hat{\Sigma}_n$ to $\Sigma_{\theta^\star}$ and then the convergence in distribution

$$\sqrt{n}\hat{\Sigma}_n^{1/2}(\widehat{\theta}_n - \theta^\star) \xrightarrow[n\to+\infty]{} \mathcal{N}_d(0, Id) \text{ under } \mathbf{P}^\star,$$

where $\mathcal{N}_d(0, Id)$ is the centered and normalised d-dimensional normal distribution. When $d = 1$, we thus consider confidence intervals of the form

$$\mathscr{I}\mathscr{C}_{\gamma,n} = \left[\widehat{\theta}_n - \frac{q_{1-\gamma/2}}{\sqrt{n}\Sigma_n^{1/2}}; \widehat{\theta}_n + \frac{q_{1-\gamma/2}}{\sqrt{n}\Sigma_n^{1/2}}\right], \tag{36}$$

where $1-\gamma$ is the asymptotic confidence level and q_z the z-th quantile of the standard normal one-dimensional distribution. In higher dimensions ($d \geq 2$), the confidence regions are more generally built relying on the chi-square distribution, namely

$$\mathscr{R}_{\gamma,n} = \{\theta \in \Theta : n\|\hat{\Sigma}_n^{1/2}(\widehat{\theta}_n - \theta)\|^2 \leq \chi_{1-\gamma}\}, \tag{37}$$

where $1-\gamma$ is still the asymptotic confidence level and now χ_z is the z-th quantile of the chi-square distribution with d degrees of freedom $\chi^2(d)$. Note that the two definitions (36) and (37) coincide when $d = 1$. Moreover, the confidence region (37) is also given by

$$\mathscr{R}_{\gamma,n} = \{\theta \in \Theta : n(\widehat{\theta}_n - \theta)^\intercal \hat{\Sigma}_n(\widehat{\theta}_n - \theta) \leq \chi_{1-\gamma}\}.$$

We present three simulation settings corresponding to the three examples developed in Section 3 and already explored in Comets et al. (2012). For each of the three simulation settings, the true parameter value $\theta^\star$ is chosen according to Table 1 and corresponds to a transient and ballistic random walk. We rely on 1000 iterations of each of the following procedures. For each setting and each iteration, we first generate a random environment according to $\nu_{\theta^\star}$ on the set of sites $\{-10^4, \dots, 10^4\}$. Note that we do not use the environment values for all the 10^4 negative sites, since only few of these sites are visited by the walk. However this extra computation cost is negligible. Then, we run a random walk in this environment and stop it successively at the hitting times T_n defined by (2), with $n \in \{10^3 k : 1 \leq k \leq 10\}$. For each stopping value n, we compute the estimators $\widehat{\theta}_n, \hat{\Sigma}_n$ and the confidence region $\mathscr{R}_{\gamma,n}$ for $\gamma = \{0.01; 0.05; 0.1\}$.

Simulation	Fixed parameter	Estimated parameter
Example I	$(a_1, a_2) = (0.4, 0.7)$	$p^\star = 0.3$
Example II	-	$(p^\star, a_1^\star, a_2^\star) = (0.3, 0.4, 0.7)$
Example III	-	$(\alpha^\star, \beta^\star) = (5, 1)$

Table 1: Parameter values for each experiment.

We first explore the convergence of $\hat{\Sigma}_n$ when n increases. We mention that the true value $\Sigma_{\theta^\star}$ is unknown even in a simulation setting (since $\tilde{\pi}_{\theta^\star}$ is unknown). Thus we can observe the convergence of $\hat{\Sigma}_n$ with n but cannot assess any bias towards the true value $\Sigma_{\theta^\star}$. The results are presented in Figures 1, 2 and 3 corresponding to the cases of Examples I, II and III, respectively. The estimators appear to converge when n increases and their variance also decreases as expected. We mention that in the cases of Examples I and II, we have 1% and 1.3%

respectively of the total $10 * 1000$ experiments for which the numerical maximisation of the likelihood did not give a result and thus for which we could not compute a confidence region.

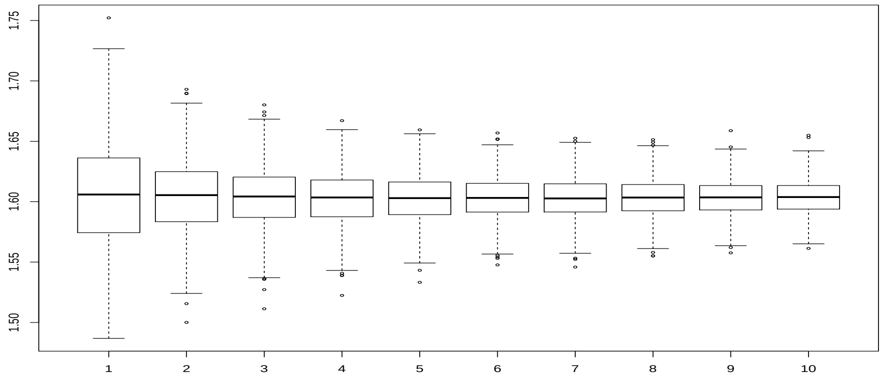

Figure 1: Boxplot of the estimator $\hat{\Sigma}_n$ obtained from 1000 iterations and for values n ranging in $\{10^3 k : 1 \leq k \leq 10\}$ in the case of Example I.

Now, we consider the empirical coverages obtained from our confidence regions $\mathscr{R}_{\gamma,n}$ in the three examples and with $\gamma \in \{0.01, 0.05, 0.1\}$ and n ranging in $\{10^3 k : 1 \leq k \leq 10\}$. The results are presented in Table 2. For the three examples, the empirical coverages are very accurate. We also note that the accuracy does not significantly change when n increases from 10^3 to 10^4. As a conclusion, we have shown that it is possible to construct accurate confidence regions for the parameter value.

	Example I			Example II			Example III		
n	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1
1000	0.994	0.952	0.899	0.992	0.953	0.909	0.977	0.942	0.901
2000	0.989	0.952	0.903	0.994	0.953	0.910	0.978	0.928	0.884
3000	0.988	0.942	0.901	0.990	0.938	0.886	0.981	0.940	0.889
4000	0.991	0.944	0.896	0.991	0.951	0.894	0.988	0.945	0.900
5000	0.990	0.942	0.896	0.993	0.942	0.891	0.986	0.941	0.883
6000	0.983	0.948	0.901	0.987	0.951	0.888	0.988	0.937	0.897
7000	0.986	0.950	0.900	0.992	0.951	0.900	0.986	0.942	0.898
8000	0.987	0.956	0.898	0.988	0.950	0.903	0.981	0.946	0.903
9000	0.990	0.959	0.913	0.990	0.949	0.893	0.985	0.939	0.901
10000	0.987	0.954	0.908	0.990	0.949	0.899	0.983	0.944	0.892

Table 2: Empirical coverages of $(1-\gamma)$ asymptotic level confidence regions, for $\gamma \in \{0.01, 0.05, 0.1\}$ and relying on 1000 iterations.

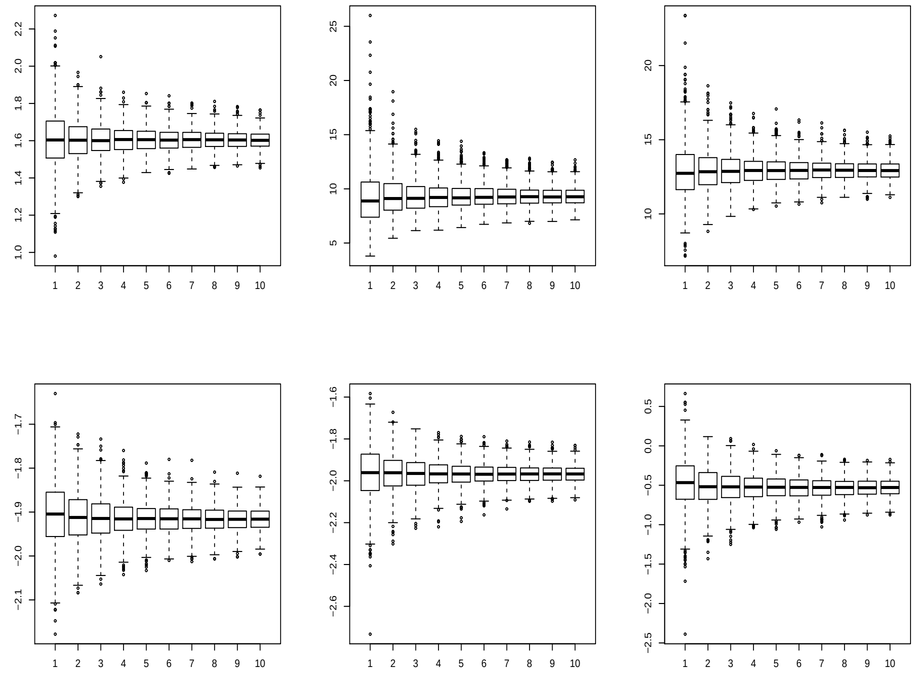

Figure 2: Boxplots of the values of the matrix $\hat{\Sigma}_n$ obtained from 1000 iterations and for values n ranging in $\{10^3 k : 1 \le k \le 10\}$ in the case of Example II. The parameter is ordered as $\theta = (\theta_1, \theta_2, \theta_3) = (p, a_1, a_2)$ and the figure displays the values: $\hat{\Sigma}_n(1,1); \hat{\Sigma}_n(2,2); \hat{\Sigma}_n(3,3); \hat{\Sigma}_n(1,2); \hat{\Sigma}_n(1,3)$ and $\hat{\Sigma}_n(2,3)$, from left to right and top to bottom.

Acknowledgments. The authors warmly thank Francis Comets and Oleg Loukianov for sharing many fruitful reflexions about this work.

References

Adelman, O. and N. Enriquez (2004). Random walks in random environment: what a single trajectory tells. *Israel J. Math. 142*, 205–220.

Benjamini, I. and H. Kesten (1996). Distinguishing sceneries by observing the scenery along a random walk path. *J. Anal. Math. 69*, 97–135.

Chernov, A. (1967). Replication of a multicomponent chain by the lightning mechanism. *Biofizika 12*, 297–301.

Comets, F., M. Falconnet, O. Loukianov, D. Loukianova, and C. Matias (2012).

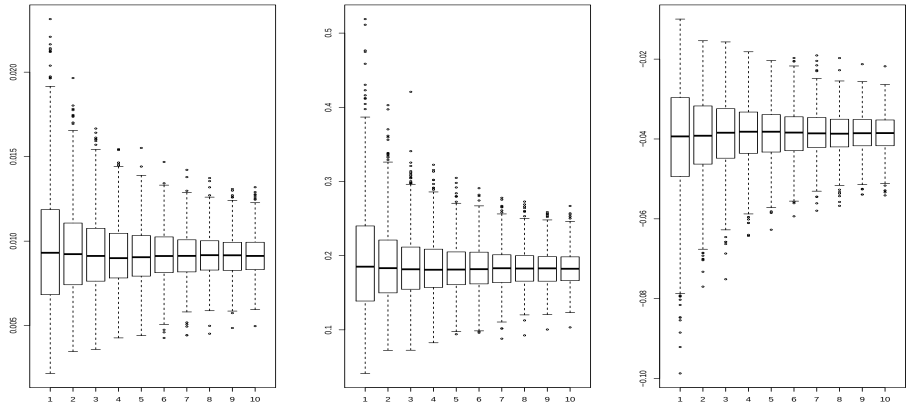

Figure 3: Boxplots of the values of the matrix $\hat{\Sigma}_n$ obtained from 1000 iterations and for values n ranging in $\{10^3 k : 1 \le k \le 10\}$ in the case of Example III. The parameter is ordered as $\theta = (\theta_1, \theta_2) = (\alpha, \beta)$ and the figure displays the values: $\hat{\Sigma}_n(1,1)$; $\hat{\Sigma}_n(2,2)$ and $\hat{\Sigma}_n(1,2)$, from left to right.

Maximum likelihood estimator consistency for ballistic random walk in a parametric random environment. Technical report, arXiv:1210.6328.

Hall, P. and C. C. Heyde (1980). *Martingale limit theory and its application.* New York: Academic Press Inc. [Harcourt Brace Jovanovich Publishers]. Probability and Mathematical Statistics.

Hughes, B. D. (1996). *Random walks and random environments. Vol. 2.* Oxford Science Publications. New York: The Clarendon Press Oxford University Press. Random environments.

Kesten, H., M. V. Kozlov, and F. Spitzer (1975). A limit law for random walk in a random environment. *Compositio Math. 30*, 145–168.

Kozlov, M. (1973). Random walk in a one-dimensional random medium. *Theory Probab. Appl. 18*, 387–388.

Löwe, M. and H. Matzinger, III (2002). Scenery reconstruction in two dimensions with many colors. *Ann. Appl. Probab. 12*(4), 1322–1347.

Matzinger, H. (1999). Reconstructing a three-color scenery by observing it along a simple random walk path. *Random Structures Algorithms 15*(2), 196–207.

Revuz, D. (1984). *Markov chains* (Second ed.), Volume 11 of *North-Holland Mathematical Library*. Amsterdam: North-Holland Publishing Co.

Solomon, F. (1975). Random walks in a random environment. *Ann. Probability 3*, 1–31.

Temkin, D. E. (1972). One-dimensional random walks in a two-component chain. *Soviet Mathematics Doklady 13*, 1172–1176.

van der Vaart, A. W. (1998). *Asymptotic statistics*, Volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge: Cambridge University Press.

Zeitouni, O. (2004). Random walks in random environment. In *Lectures on probability theory and statistics*, Volume 1837 of *Lecture Notes in Math.*, pp. 189–312. Berlin: Springer.